PART II: HELPER T-CELL EPITOPES

SUMMARY

Part II includes tables, maps, and alignments of HIV-specific specific helper T-cell (Th) epitopes arranged sequentially according to the location of the proteins in the HIV-1 genome. We attempted to make this section as comprehensive as possible, requiring that the epitope be contained within a region of approximately 30 amino acids maximum, but not that the precise boundaries be defined. The HLA specificity is usually not determined for Th epitopes. For more recent updates and useful searching capabilities, please see our Web site: http://hiv-web.lanl.gov/immuno. The same epitope can have multiple entries, as each entry represents a single publication.

TABLES:

Each Th epitope has a six-part basic entry:

- **Location:** The amino acid positions of the epitope boundaries and the reference sequence are listed as given in the primary publication. Frequently, these positions as published are imprecise, and do not truly correspond to the numbering of the sequence, but they provide a reasonable guide to the peptide's approximate location in the protein. Also, in many cases the reference sequence identification was not provided, and in such cases it is not possible to use these numbers to specify precise locations.
- **WEAU Location:** The viral strain WEAU (GenBank Accession Number U21135) is used as a reference strain throughout this publication. The position of the defined epitope location on the sequence of the WEAU protein is indicated. Obviously WEAU may not be identical to a given defined reactive sequence, so we simply indicating the location of the aligned positions. The WEAU numbering is used in to the protein maps in this database. Nef in the WEAU cloned sequence has a frame shift, but the Nef reference protein sequence was completed past the frame shift stop codon for the purpose of mapping the epitope locations.
- **Epitope:** The amino acid sequence of the epitope of interest as defined in the reference, based on the reference strain used in the study defining the epitope. On rare occasions, when only the epitope location and not the actual epitope was specified in the original publication, and the sequences were numbered inaccurately by the primary authors, we may have misrepresented the epitope's amino acid sequence. Therefore, epitopes that were not explicitly written out in the text in the primary publication, those that we determined by looking up the reference strain and the numbered location, are followed by a question mark in the table.
- **Antigen:** The antigenic stimulus of the Th response to the defined epitope.
- **Species(HLA):** The species responding and HLA specificity of the epitope, when known.
- **Reference** The primary reference.

Following each entry for a given Th epitope is a brief comment explaining the context of the study that defined the epitope. If the same epitope was studied in several labs, each study is cited in its own entry.

HIV PROTEIN EPITOPE MAPS:

All human and primate Th epitopes defined to within 21 amino acids or less are indicated on the HIV protein epitope maps. HLA restriction information is included when known.

The location and HLA restriction elements (when known) of Th epitopes are indicated on protein sequences of the WEAU clone 1.60. These maps are meant to provide the relative location of epitopes on a given protein, but the WEAU sequence may not actually carry the epitope of interest, as it may vary relative to the sequence for which the epitope was defined. Epitopes are numbered, and the numbering on this map is used to reference the corresponding epitope sequence alignments. The WEAU sequence is described on page I-1.

ALIGNMENTS:

As with the maps, only human and primate Th epitopes defined to within less than or equal to 21 amino acids, with a known HLA specificity, have corresponding alignments. For each numbered epitope in the epitope-protein maps, an alignment was generated from the protein sequence alignments in the HIV-1 genetic sequence database. All epitopes are aligned to the subtype B consensus (the most common amino acid found in subtype B in each position), with the sequence used to defined the epitope indicated directly beneath the B consensus. In consensus sequences an upper case letter indicates the amino acid was present in all sequences, a lower case letter indicates the amino acid was present in most sequences in a given position, and a question mark indicates two or more amino acids were represented with equal frequency. In the master alignment files from which the epitope alignments were created, there are many partial sequences that do not span entire genes. Many of these partial sequences were removed before extracting the epitope alignments, but after the consensus sequences were calculated. As a result, the consensuses as shown in the epitope alignments correctly record the consensus of the master alignment of the database but not necessarily the consensus of the sequences shown in the epitope alignment. We used the full length proteins in the 1997 HIV-1 database protein alignments for this section. The alignments were modified to optimize the alignment relative to the defined epitope and minimize insertions and deletions. A dash indicates identity to the consensus sequence, and a period indicates an insertion made to maintain the alignment. Stop codons are indicated with a $, and frameshifts by a #; some epitopes are only partially sequenced.

REFERENCES AND NOTES